MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Equivalence of Topologies Spanned by the Knothe-Rosenblatt and the Adapted Wasserstein Distance

verfasst von / submitted by

Alexander Posch, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Univ.-Prof. Dr. Mathias Beiglböck

Abstract

In this master's thesis, we extend the Knothe-Rosenblatt coupling [24, p. 20] to probability measures on arbitrary Polish metric spaces. We use such couplings to define the p-Knothe-Rosenblatt distance and show that this distance forms a metric on $\mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$. However, we show with an example that this does not hold for arbitrary spaces because the p-Knothe-Rosenblatt distance does not necessarily satisfy the triangle inequality. In Theorem 3.4, the main result, we show that the p-Knothe-Rosenblatt distance spans the same topology as the adapted p-Wasserstein distance. This result is of interest since the p-Knothe-Rosenblatt distance is often simpler to calculate than the adapted p-Wasserstein distance. The theorem extends the statement in [4], where the equivalence of four further approaches for adapted Wasserstein topologies is demonstrated. Our proof is by induction over the time steps. Finally, we show that for $\mathcal{X}$ compact $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$, the space of stochastic processes on $\mathcal{X}$ with $n > 1$ time steps equipped with the p-Knothe-Rosenblatt distance is not necessarily precompact. In contrast $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ equipped with the adapted p-Wasserstein distance is precompact.

Zusammenfassung

In dieser Masterarbeit erweitern wir die Knothe-Rosenblatt Kopplung [24, p. 20] auf Wahrscheinlichkeitsmaße auf beliebigen polnischen metrischen Räumen. Wir nutzen diese Knothe-Rosenblatt Kopplungen zur Definition der p-Knothe-Rosenblatt Distanz und zeigen, dass diese Distanz eine Metrik auf $\mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$ bildet. Für beliebige Räume ist dies jedoch nicht der Fall. Wir zeigen anhand eines Beispiels, dass die Dreiecksungleichung im Allgemeinen nicht immer erfüllt ist. In Theorem 3.4, dem Hauptresultat, zeigen wir dass die Knothe-Rosenblatt Distanz dieselbe Topologie wie die adaptierte Wasserstein Distanz aufspannt. Dieses Resultat ist von Interesse, da die p-Knothe-Rosenblatt Distanz oft leichter zu berechnen ist als die adaptierte Wasserstein Distanz. Das Theorem erweitert das Statement in [4] in dem die Äquivalenz vier weiterer Ansätze die adaptierte Wasserstein Topologien zu definieren gezeigt wird. Unser Beweis erfolgt durch Induktion über die Zeitschritte. Schließlich zeigen wir, dass, für $\mathcal{X}$ kompakt, $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$, der Raum der stochastischen Prozesse auf $\mathcal{X}$ mit $n > 1$ Zeitschritten, ausgestattet mit der p-Knothe-Rosenblatt Distanz, nicht notwendigerweise präkompakt ist. $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$, ausgestattet mit der adaptierten p-Wasserstein Distanz, ist dagegen präkompakt.

Acknowledgement

I would like to thank my advisor, Mathias Beiglböck, for introducing me to the fascinating field of Stochastic Optimal Transport, suggesting the topic of this thesis, and for providing invaluable guidance and support. Special thanks also to Gudmund Pammer for co-superving this thesis and for sharing his insights in the formulation of the proofs. Last but not least, I wish to express my thanks to my family and friends for their continuous support throughout my studies.

Contents

1. Introduction

Discrete time stochastic processes with $d \in \mathbb{N}$ time steps on a Polish metric space $\mathcal{X}$ can be represented as probability measures on $\mathcal{X}^n$. A common approach to establish a topology on $\mathcal{P}(\mathcal{X}^n, \| \,.\, \|_p)$ is the standard weak topology. A sequence of measures $(\mu_i)_{i\geq 0}$ is said to converge weakly to a measure μ if for all bounded Lipschitz continuous functions f we have that $\lim_{i\to\infty} \int f \, d\mu_i = \int f \, d\mu$.

The weak topology is induced by the p-Wasserstein distance for $p \geq 1$. The Wasserstein distance represents the cost needed to move masses between two distributions. This cost takes into account the amount of mass being moved and the distance it is moved. For a formal definition, see Definition 2.2. The Wasserstein distance itself has many applications, for example, in machine learning [19, 11], meteorology [9] and computer vision [14].

Using the Wasserstein distance to compare stochastic processes embedded in $\mathcal{X}^n$, however, does not take the time dependent structure of the stochastic process into account. Therefore, the weak topology cannot ensure continuity for optimal stopping, the Doob-decomposition, and stochastic control problems. Continuity is important for many applications, for example in optimal stopping problems, where, even if the the exact distribution of a stochastic process is not known, we still want the solutions to be approximately correct.

To address these issues, a number of different approaches to define distances and topologies on the space of stochastic processes were introduced [15, 1, 8, 18, 12]. This includes a so-called information topology for tackling sequential decision problems [15], a variation of the weak topology for prediction processes [1], and three variations of the Wasserstein distance [8, 12, 18], which can be used to establish topologies. It was shown in [4] that these five approaches yield the same topology and that this topology is the coarsest one that guarantees continuity for optimal stopping problems with bounded continuous rewards. One approach for establishing this topology is defining an adapted version of the Wasserstein distance, see Definition 2.4.

In this master's thesis, we show that yet another approach yields the same topology as the five given above. In particular, we show in Theorem 3.4 that the topology induced by the Knothe-Rosenblatt distance defined in Definition 2.4 below, coincides with the topology induced by the adapted Wasserstein distance.

In addition we prove some properties of the Knothe-Rosenblatt distance, whether it is a metric and discuss precompactness.

The Knothe-Rosenblatt coupling was first introduced by Knothe and Rosenblatt in their respective works [17, 23]. In [23] Rosenblatt introduces the coupling and discusses some properties and possible applications. Independently, Knothe defines and explores this coupling in [17], examining generalizations of the Brunn-Minkowski and the isoperimetric inequality.

Adapted versions of the Wasserstein distance find applications in many areas in mathematical finance. In [21, 22], a nested distance is introduced, showcasing its utility in multistage stochastic optimization problems by using it for judging how similar stochastic processes are to approximations. Also, in [16], they approximate stochastic processes to solve an optimization problem and use the nested distance to quantify the fit of the approximations, and in [5], stability of solutions to optimal stopping problems and multi-period stochastic optimization problems with regard to the adapted Wasserstein distance are analyzed.

Also, the adapted Wasserstein topology has multiple applications. For example, in [13], the adapted Wasserstein Topology is used to define confidence sets for estimated models for asset price processes. This is then used to address a stochastic optimization problem under model uncertainty. In [3], robustness of optimal hedging strategies with regard to the adapted Wasserstein topology is shown, and in [6], approximations of martingale couplings in the adapted weak topology are explored.

This master thesis is structured as follows. In Section 2, we introduce all the necessary notation and definitions. In Section 3, we prove Theorem 3.4, the main result of this thesis and necessary auxiliary lemmas. In Section 4, we prove that on $\mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$ the triangle inequality holds for the Knothe-Rosenblatt distance. We also give a counterexample showing that the Knothe-Rosenblatt distance does not satisfy the triangle inequality for probability measures defined on arbitrary Polish metric spaces. In Section 5, we show that, in contrast to the adapted Wasserstein distance [4, Theorem 1.7], the completion of the weak adapted topology with regard to the Knothe-Rosenblatt distance is not precompact on subsets where the laws are tight and uniformly integrable.

2. Definitions and notation

We consider a Polish metric space $(\mathcal{X}, \rho_{\mathcal{X}})$ and for $n \in \mathbb{N}$ the product spaces $\mathcal{X}^n = \prod_{i=1}^n \mathcal{X}_i$, where $\mathcal{X}_i = \mathcal{X}$ for all i, together with the metric

$$\rho_{\mathcal{X}^n,p}((x_t)_{t=1}^n, (y_t)_{t=1}^n) = \left(\sum_{t=1}^n \rho_{\mathcal{X}}(x_t, y_t)^p \right)^{\frac{1}{p}}$$

where $p \geq 1$. We use $(X_t)_{t=1}^n$ to denote the canonical process on $\mathcal{X}^n$, i.e. X_t is the projection onto $\mathcal{X}_t$. On $\mathcal{X}^n \times \mathcal{X}^n$ call $X = (X_t)_{t=1}^n$ the projection on the first factor and $Y = (Y_t)_{t=1}^n$ the projection on the second factor. Given $k \in \{1, \ldots, n\}$ and a measure η on $\mathcal{X}^n$ we set

$$\eta_k := \mathrm{proj}_{\prod_{i=1}^k \mathcal{X}_i} \eta$$

and for a measure π on $\mathcal{X}^n \times \mathcal{X}^n$ we set

$$\pi_k := \mathrm{proj}_{(\prod_{i=1}^k \mathcal{X}_i) \times (\prod_{i=1}^k \mathcal{X}_i)} \pi.$$

Furthermore, for $\boldsymbol{x} \in \mathcal{X}^k$ we let $\eta^{\boldsymbol{x}}$ denote the disintegration of the measure η with regard to the first k variables at $\boldsymbol{x}$, which exists and is η_k almost everywhere unique due to the disintegration theorem.

Definition 2.1. [4, p. 3] *For measures μ and ν on $\mathcal{X}^n$ the set of couplings from μ to ν is defined by*

$$\mathrm{Cpl}(\mu, \nu) = \{\pi \in \mathbb{P}(\mathcal{X}^n \times \mathcal{X}^n) | X \sim \mu \text{ and } Y \sim \nu \text{ under } \pi\}.$$

We say a coupling π is causal if

$$\pi((Y_1, \ldots, Y_k) \in A | X) = \pi((Y_1, \ldots, Y_k) \in A | X_1, \ldots, X_k), \forall k \leq n$$

and all measurable $A \subseteq \mathcal{X}^k$. We say a coupling π is bicausal if the condition also holds with the roles of X and Y exchanged. We write $\mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$ for the set of all bicausal couplings from μ to ν. A coupling $\gamma \in \mathrm{Cpl}(\mu, \nu)$ is considered optimal (cost minimizing) for the p-Wasserstein distance ($p \geq 1$), if

$$\left(\int \rho_{\mathcal{X}^n,p}^p \, d\gamma \right)^{\frac{1}{p}} = W_p(\mu, \nu) := \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \left(\int \rho_{\mathcal{X}^n,p}^p \, d\pi \right)^{\frac{1}{p}}$$

and we call $W_p(\mu, \nu)$ the p-Wasserstein distance between μ and ν

The next definition extends the definition of the Knothe-Rosenblatt coupling as described in [24, p. 20] to spaces where optimal couplings are not necessarily unique.

Definition 2.2. *Let μ and ν be probability measures on the Polish metric space $(\mathcal{X}^n, \rho_{\mathcal{X}^n,p})$. We define the set of all p-Knothe-Rosenblatt couplings from μ to ν, $\mathrm{Cpl}_{\mathrm{KR}p}(\mu,\nu)$, as the set of all measures π on $\mathcal{X}^n \times \mathcal{X}^n$ that satisfy that π_1 is an optimal coupling from μ_1 to ν_1 and for all $2 \leq k \leq n$, $\pi_k^{\boldsymbol{x},\boldsymbol{y}}$ is an optimal coupling from $\mu_k^{\boldsymbol{x}}$ to $\nu_k^{\boldsymbol{y}}$ with regard to the p-Wasserstein distance, for π_{k-1} almost all $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^{k-1} \times \mathcal{X}^{k-1}$. Here $\pi_k^{\boldsymbol{x},\boldsymbol{y}}$ denotes the disintegration of π_k at $((x_1,y_1),\dots,(x_{k-1},y_{k-1}))$.*

Remark 2.3. *It can be show by a standard measurable selection argument that for any μ and ν, probability measures on the Polish metric space $(\mathcal{X}^n, \rho_{\mathcal{X}^n,p})$, the set $\mathrm{Cpl}_{\mathrm{KR}p}(\mu,\nu)$ is non-empty. Furthermore, from the definitions it follows that for μ and ν, measures on $\mathcal{X}^n \times \mathcal{X}^n$, and $k < n$ we have $\{\pi_k | \pi \in \mathrm{Cpl}_{\mathrm{KR}p}(\mu,\nu)\} = \mathrm{Cpl}_{\mathrm{KR}p}(\mu_k,\nu_k)$ and $\{\pi_k | \pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu,\nu)\} = \mathrm{Cpl}_{\mathrm{BC}}(\mu_k,\nu_k)$.*

Definition 2.4. *For μ and ν measures on the Polish metric space $(\mathcal{X}^n, \rho_{\mathcal{X}^n,p})$ and $p \geq 1$ we call*

$$AW_p(\mu,\nu) := \inf_{\pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu,\nu)} \left(\int \rho^p_{\mathcal{X}^n,p} \, d\pi \right)^{\frac{1}{p}}$$

the adapted p-Wasserstein distance from μ to ν, as in [4]. Additionally we define

$$KR_p(\mu,\nu) := \sup_{\pi \in \mathrm{Cpl}_{\mathrm{KR}p}(\mu,\nu)} \left(\int \rho^p_{\mathcal{X}^n,p} \, d\pi \right)^{\frac{1}{p}}$$

as the p-Knothe-Rosenblatt distance from μ to ν.

We can compose couplings by first transporting the mass from the first space to the second and then from the second to the third. This results in a transport from the first space to the third space. For Monge transports [24] this composition is given by the function composition of the two transport maps. To expand this concept to non deterministic transport plans, we use the notion of the gluing of couplings.

Definition 2.5. [24, p. 23 (Gluing Lemma)] *Let μ,ν,η be measures on the Polish spaces $\mathcal{X},\mathcal{Y},\mathcal{Z}$. Given $\pi \in \mathrm{Cpl}(\mu,\nu)$ and $\gamma \in \mathrm{Cpl}(\nu,\eta)$ we define $\pi \underset{\mathcal{Y}}{\otimes} \gamma$ as the measure on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ such that*

$$\pi \underset{\mathcal{Y}}{\otimes} \gamma(dx_1,dx_2,dx_3) = \pi(dx_1,dx_2)\gamma^{x_2}(dx_3).$$

Thus $proj_{\mathcal{X},\mathcal{Y}}(\pi \underset{\mathcal{Y}}{\otimes} \gamma) = \pi$*,* $proj_{\mathcal{Y},\mathcal{Z}}(\pi \underset{\mathcal{Y}}{\otimes} \gamma) = \gamma$ *and we call*

$$\pi \circledast \gamma := proj_{\mathcal{X},\mathcal{Z}}(\pi \underset{\mathcal{Y}}{\otimes} \gamma) \in \mathrm{Cpl}(\mu, \eta)$$

the marriage of π *and* γ*.*

Another useful tool is the modulus of continuity. It is a measure of how sensitive a measure in two variables is to perturbations of its variables.

Definition 2.6. [10, Definition 1.2 (Modulus of Continuity)] *Let* $p \geq 1$*,* $(\mathcal{X}, \rho_{\mathcal{X}}), (\mathcal{Y}, \rho_{\mathcal{Y}})$ *be Polish metric spaces and let* μ *be a probability measure on* $\mathcal{X} \times \mathcal{Y}$*. Then we define the modulus of continuity* $\omega_{\mu,p} : \mathbb{R}_+ \to \mathbb{R}_+$ *of* μ *by*

$$\omega_{\mu,p}(\delta) := \sup_{\gamma \in Per(\mu,\delta)} \left(\int \rho_{\mathcal{Y}}^p \, d\gamma \right)^{\frac{1}{p}}$$

where $Per(\mu, \delta) := \{\gamma \in \mathrm{Cpl}(\mu, \mu) | \left(\int \rho_{\mathcal{X}}^p \, d\gamma\right)^{\frac{1}{p}} \leq \delta\}$*. Thus any coupling in* $\mathrm{Cpl}(\mu, \mu)$ *with a cost in* $\mathcal{X}$ *that is less or equal to* δ *has a cost in* $\mathcal{Y}$ *that is at most* $\omega_{\mu,p}(\delta)$*. An important property of the modulus of continuity is* [10, Lemma 2.7]*. It states that the modulus of continuity* $\omega_{\mu,p}$ *is continuous at* 0 *if and only if the measure* μ *is concentrated on the graph of a function.*

Definition 2.7. [10, p. 2] *For probability measures* ν *on the Polish space* $\mathcal{X}^n$ *we define*

$$\mathcal{I}(\nu) := ((\boldsymbol{x} \to (\boldsymbol{x}, \nu^{\boldsymbol{x}}))_{\#}\nu_{n-1}$$

for $\boldsymbol{x} \in \mathcal{X}^{n-1}$*. This is well defined since* $x \to \nu^x$ *is a measurable function by the disintegration theorem. Thus* $\mathcal{I}(\nu)$ *is a probability measure on* $\mathcal{X}^{n-1} \times \mathbb{P}(\mathcal{X})$ *and is by definition concentrated on the graph of a function.*

Our goal is to establish Theorem 3.4 stating that the p-Knothe-Rosenblatt distance induces the p weak adapted topology. We start by proving a few auxiliary lemmas.

3. The topology induced by Knothe-Rosenblatt is equivalent to the one induced by adapted Wasserstein

In this section, we prove Theorem 3.4, the main result of this thesis. We start by proving two auxiliary lemmas. Lemma 3.1 and Lemma 3.2 give equivalent definitions of bicausal couplings. Lemma 3.2 also allows us to construct bicausal couplings. Both lemmas are a direct consequence of [2, Proposition 5.1] or [25, Proposition 2.2.7] but we prove them here directly.

Lemma 3.1. *For measures μ, ν on $\mathcal{X}^n$ and $\pi \in \mathrm{Cpl}(\mu, \nu)$ the following statements are equivalent:*

(i) $\pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$
(ii) $\forall k \leq n$ *we have* $\pi^{\boldsymbol{x},\boldsymbol{y}} \in \mathrm{Cpl}(\mu^{\boldsymbol{x}}, \nu^{\boldsymbol{y}})$ *for* π_k *almost all* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}^k$.

Proof. For ease of notation we will write X_a^b for $(X_a, X_{a+1}, \ldots, X_b)$.

To prove this statement, we will first show that the two statements (i') and (ii'), defined below, are equivalent. Then we show that (i), (ii) in Lemma 3.1 are equivalent to the statements (i'), (ii'), respectively.

Let statement (i') be that,

$$\pi(Y_1^k \in A | X_{k+1}^n \in B \wedge X_1^k = x_1^k) = \pi(Y_1^k \in A | X_1^k = x_1^k) \tag{i'}$$

for all $k \leq n$, all measurable $A \subseteq \mathcal{X}^k$, $B \subseteq \mathcal{X}^{n-k}$ with $\pi(Y_1^k \in A | X_1^k) > 0$ and $\pi(X_{k+1}^n \in B | X_1^k) > 0$ and π_k almost all x_1^k. Furthermore this equation also holds with the roles of X and Y exchanged.

Let statement (ii') be that,

$$\pi(X_{k+1}^n \in B | Y_1^k \in A \wedge X_1^k = x_1^k) = \pi(X_{k+1}^n \in B | X_1^k = x_1^k) \tag{ii'}$$

for all $k \leq n$, all measurable $A \subseteq \mathcal{X}^k$, $B \subseteq \mathcal{X}^{n-k}$ with $\pi(Y_1^k \in A | X_1^k) > 0$ and $\pi(X_{k+1}^n \in B | X_1^k) > 0$ and π_k almost all x_1^k. Furthermore this equation holds with the roles of X and Y exchanged.

Rewriting statement (ii'), we get

$$\begin{aligned}
&\frac{\pi(X_{k+1}^n \in B \wedge Y_1^k \in A | X_1^k = x_1^k)}{\pi(Y_1^k \in A | X_1^k = x_1^k)} = \pi(X_{k+1}^n \in B | X_1^k = x_1^k) \\
\Leftrightarrow\ &\frac{\pi(X_{k+1}^n \in B \wedge Y_1^k \in A | X_1^k = x_1^k)}{\pi(X_{k+1}^n \in B | X_1^k = x_1^k)} = \pi(Y_1^k \in A | X_1^k = x_1^k) \\
\Leftrightarrow\ &\pi(Y_1^k \in A | X_{k+1}^n \in B \wedge X_1^k = x_1^k) = \pi(Y_1^k \in A | X_1^k = x_1^k)
\end{aligned}$$

for all π_k almost all x_1^k and all measurable $A \subseteq \mathcal{X}^k$, $B \subseteq \mathcal{X}^{n-k}$ with $\pi(Y_1^k \in A | X_1^k = x_1^k) > 0$, $\pi(X_{k+1}^n \in B | X_1^k = x_1^k) > 0$. The same also holds with the roles of X and Y exchanged. Thus, statements (i') and (ii') are equivalent.

We will now show that the statements (i) and (i') are equivalent, the proof for why statements (ii) and (ii') are equivalent will follow analogously.

To see that (i) implies (i') we simply observe that $x_1^k \times B$ is measurable and thus (i') follows directly by (i).

To show that, (i') implies (i) we will show that for π_k almost all x_1^k and $\pi^{x_1^k}$ almost all x_{k+1}^n we have

$$\pi(Y_1^k \in A | X_{k+1}^n = x_{k+1}^n \wedge X_1^k = x_1^k) = \pi(Y_1^k \in A | X_1^k = x_1^k) \tag{3.1}$$

which in turn implies statement (i).

First, we note that statement (i') also holds for all A with $\pi(Y_1^k \in A | X_1^k = x_1^k) = 0$ since then $\pi(Y_1^k \in A^c | X_1^k = x_1^k) = 1$ and thus

$$\begin{aligned}
&\pi(Y_1^k \in A | X_{k+1}^n \in B \wedge X_1^k = x_1^k) \\
&= 1 - \pi(Y_1^k \in A^c | X_{k+1}^n \in B \wedge X_1^k = x_1^k) \\
&= 1 - \pi(Y_1^k \in A^c | X_1^k = x_1^k) \\
&= \pi(Y_1^k \in A | X_1^k = x_1^k).
\end{aligned}$$

This means that the measures coincide for all sets A. We will now prove equation (3.1) by considering a sequence of sets that converges to $\{x_{k+1}^n\}$ from above.

First, observe that since X^{n-k} is separable, for $\pi(\,\cdot\,| X_1^k = x_1^k)$ almost all $x_{k+1}^n \in X^{n-k}$ and all $\varepsilon > 0$ we have $\pi(B_\varepsilon(x_{k+1}^n) | X_1^k = x_1^k) > 0$, where $B_\varepsilon(x)$ is the open ball around x with radius ε. Thus, for almost all x_{k+1}^n, and all sequences of open sets $(C_{n,x_{k+1}^n})_{n\geq 1} \searrow \{x_{k+1}^n\}$

$$\frac{\pi(Y_1^k \in A \wedge X_{k+1}^n \in C_{n,x_{k+1}^n} | X_1^k = x_1^k)}{\pi(X_{k+1}^n \in C_{n,x_{k+1}^n} | X_1^k = x_1^k)} = \pi(Y_1^k \in A | X_{k+1}^n \in C_{n,x_{k+1}^n} \wedge X_1^k = x_1^k)$$

is well defined. Furthermore, since

$$\pi(Y_1^k \in A | X_{k+1}^n \in C_{n,x_{k+1}^n} \wedge X_1^k = x_1^k) = \pi(Y_1^k \in A | X_1^k = x_1^k)$$

the sequence of probability measures $\pi(Y_1^k \in \cdot | X_{k+1}^n \in C_{n,x_{k+1}^n} \wedge X_1^k = x_1^k)$ is not only convergent but actually constant for almost all x_{k+1}^n. Now, by [20] for almost all x_{k+1}^n this sequence weakly converges to $\pi(Y_1^k \in \cdot | X_{k+1}^n = x_{k+1}^n \wedge X_1^k = x_1^k)$ and since the sequence is constant we also have setwise convergence. Therefore,

$$\begin{aligned}
&\pi(Y_1^k \in A | X_{k+1}^n = x_{k+1}^n \wedge X_1^k = x_1^k) \\
&= \lim_{n\to\infty} \pi(Y_1^k \in A | X_{k+1}^n \in C_{n,x_{k+1}^n} \wedge X_1^k = x_1^k) \\
&= \lim_{n\to\infty} \pi(Y_1^k \in A | X_1^k = x_1^k) \\
&= \pi(Y_1^k \in A | X_1^k = x_1^k),
\end{aligned}$$

which is (i') and what we wanted to show.

Next we show that the statements (ii) and (ii') are equivalent. As stated this proof will be very similar to $(i) \Leftrightarrow (i')$.

To see that (ii) implies (ii'), we simply integrate $\pi(X_{k+1}^n \in B | X_1^k = x_1^k, Y = y_1^k)$ over all $y_1^k \in A$, then since

$$\pi(X_{k+1}^n \in B | X_1^k = x_1^k, Y_1^k = y_1^k) = \pi(X_{k+1}^n \in B | X_1^k = x_1^k)$$

holds for almost all y_1^k we also have

$$\pi(X_{k+1}^n \in B | X_1^k = x_1^k, Y_1^k \in A) = \pi(X_{k+1}^n \in B | X_1^k = x_1^k),$$

which is statement (ii').

For the other direction, (ii') implies (ii), we will show that for π_k almost all x_1^k and $\pi^{x_1^k}$ almost all y_1^k we have

$$\pi(X_{k+1}^n \in B | Y_1^k = y_1^k \wedge X_1^k = x_1^k) = \pi(X_{k+1}^n \in B | X_1^k = x_1^k) \tag{3.2}$$

which is equivalent to statement (ii). Here note that the integral is well defined since by the disintegration theorem $y_1^k \to \pi(X_{k+1}^n \in B | Y_1^k = y_1^k \wedge X_1^k = x_1^k)$ is a measurable function.

First, we again note that statement (i') also holds for all B with $\pi(X_{k+1}^n \in B | X_1^k = x_1^k) = 0$ since then $\pi(X_{k+1}^n \in B^c | X_1^k = x_1^k) = 1$ and thus

$$\begin{aligned}
&\pi(X_{k+1}^n \in B | Y_1^n \in A \wedge X_1^k = x_1^k) \\
&= 1 - \pi(X_{k+1}^n \in B^c | Y_1^k \in A \wedge X_1^k = x_1^k) \\
&= 1 - \pi(X_{k+1}^n \in B^c | X_1^k = x_1^k) \\
&= \pi(X_{k+1}^n \in B | X_1^k = x_1^k).
\end{aligned}$$

This means that the measures coincide for all sets B. We will now prove equation (3.2) by considering a sequence of sets that converges to $\{y_1^k\}$ from above.

First we again observe that since $\mathcal{X}^k$ is separable, for $\pi_k(\,\cdot\,| X_1^k = x_1^k)$ almost all $y_1^k \in \mathcal{X}^k$ and all $\varepsilon > 0$ we have $\pi(B_\varepsilon(y_1^k) | X_1^k = x_1^k) > 0$, where $B_\varepsilon(x)$ is the open ball around x with radius ε. Thus, for almost all y_1^k and all sequences of open sets $(C_{n,y_1^k})_{n\geq 1} \searrow \{y_1^k\}$

$$\frac{\pi(Y_1^k \in A \wedge X_{k+1}^n \in C_{n,x_{k+1}^n} | X_1^k = x_1^k)}{\pi(X_{k+1}^n \in C_{n,x_{k+1}^n} | X_1^k = x_1^k)} = \pi(Y_1^k \in A | X_{k+1}^n \in C_{n,x_{k+1}^n} \wedge X_1^k = x_1^k)$$

is well defined. Furthermore, since

$$\pi(X_{k+1}^n \in B | Y_1^k \in C_{n,y_1^k} \wedge X_1^k = x_1^k) = \pi(X_{k+1}^n \in B | X_1^k = x_1^k)$$

the sequence of probability measures $\pi(X_{k+1}^n \cdot | X_{k+1}^n \in C_{n,y_1^k} \wedge X_1^k = x_1^k)$ is not only convergent but actually constant for almost all y_1^k. Now, by [20] for almost all y_1^k this sequence weakly converges to $\pi(X_{k+1}^n \,\cdot\, | Y_1^k = y_1^k \wedge X_1^k = x_1^k)$ and since the sequence is constant we also have setwise convergence. Therefore,

$$\begin{aligned}
&\pi(X_{k+1}^n \in B | Y_1^k = y_1^k \wedge X_1^k = x_1^k) \\
&= \lim_{n\to\infty} \pi(X_{k+1}^n \in B | Y_1^k \in C_{n,y_1^k} \wedge X_1^k = x_1^k) \\
&= \lim_{n\to\infty} \pi(X_{k+1}^n \in B | X_1^k = x_1^k) \\
&= \pi(X_{k+1}^n \in B | X_1^k = x_1^k),
\end{aligned}$$

which is (ii') and what we wanted to show. $\square$

Lemma 3.2. *Let $\mu, \nu \in \mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$, $0 \le n' \le n$ and $\pi \in \mathcal{P}(\mathcal{X}^n \times \mathcal{X}^n)$. Then the following statements are equivalent:*

(i) $\pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$.

(ii) $\pi_{n'} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_{n'}, \nu_{n'})$ *and* $\pi^{x,y} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu^x, \nu^y)$ *for* $\pi_{n'}$ *almost all* $(x, y) \in \mathcal{X}^{n'} \times \mathcal{X}^{n'}$.

Proof. For the first direction, take $\pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$. Clearly, we have that $\pi_{n'} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_{n'}, \nu_{n'})$. To show that $\pi^{x_1^{n'}, y_1^{n'}} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu^{x_1^{n'}}, \nu^{y_1^{n'}})$ for $\pi_{n'}$ almost all $(x_1^{n'}, y_1^{n'}) \in \mathcal{X}^{n'} \times \mathcal{X}^{n'}$ we use Lemma 3.1 and see that indeed for all $k > n'$ and $(x_{n'+1}^k, y_{n'+1}^k) \in \mathcal{X}^{k-n'} \times \mathcal{X}^{k-n'}$ it holds that

$$(\pi^{x_1^{n'}, y_1^{n'}})^{x_{n'+1}^k, y_{n'+1}^k} = \pi^{x_1^k, y_1^k} \in \mathrm{Cpl}(\mu^{x_1^k}, \nu^{y_1^k}) = \mathrm{Cpl}((\mu^{x_1^{n'}})^{x_{n'+1}^k}, (\nu^{y_1^{n'}})^{y_{n'+1}^k}).$$

For the other direction, we also use Lemma 3.1. Take a measure π on $\mathcal{X}^n \times X^n$ with $\pi_{n'} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_{n'}, \nu_{n'})$ and $\pi^{x,y} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu^x, \nu^y)$ for $\pi_{n'}$ almost all $(x, y) \in \mathcal{X}^{n'} \times \mathcal{X}^{n'}$. We want to show that for $k \le n$ and π_k almost all $(x, y) \in \mathcal{X}^k \times X^k$ we have $\pi^{x,y} \in \mathrm{Cpl}(\mu^x, \nu^y)$. For $k = n'$ this condition is naturally fulfilled by assumption. Next, we assume $k > n'$ and $(x_1^k, y_1^k) \in \mathcal{X}^k \times \mathcal{X}^k$. Then we have $\pi^{x_1^k, y_1^k} = (\pi^{x_1^{n'}, y_1^{n'}})^{x_{n'+1}^k, y_{n'+1}^k}$ and since by our assumption $\pi^{x_1^{n'}, y_1^{n'}} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu^{x_1^{n'}}, \nu^{y_1^{n'}})$, by Lemma 3.1, we have

$$\pi^{x_1^k, y_1^k} = (\pi^{x_1^{n'}, y_1^{n'}})^{x_{n'+1}^k, y_{n'+1}^k} \in \mathrm{Cpl}((\mu^{x_1^{n'}})^{x_{n'+1}^k}, (\nu^{y_1^{n'}})^{y_{n'+1}^k}) = \mathrm{Cpl}(\mu^{x_1^k}, \nu^{y_1^k}).$$

Finally, we assume that $k < n'$ and $(x_1^k, y_1^k) \in \mathcal{X}^k \times \mathcal{X}^k$ and $A \subseteq X^{n-k}$ measurable. Then

$$\begin{aligned}
&\pi^{x_1^k, y_1^k}(x_{k+1}^n \in A) \\
&= \int 1_A(x_{k+1}^n)\, d\pi^{x_1^k, y_1^k}(x_{k+1}^n, y_{n'+1}^n) \\
&= \int 1_A(x_{k+1}^n)\, d(\pi^{x_1^k, y_1^k})^{x_{k+1}^{n'}, y_{k+1}^{n'}}(x_{n'+1}^n, y_{n'+1}^n)\, d\pi_{n'}^{x_1^k, y_1^k}(x_{k+1}^{n'}, y_{k+1}^{n'}) \\
&= \int 1_A(x_{k+1}^n)\, d\pi^{x_1^{n'}, y_1^{n'}}(x_{n'+1}^n, y_{n'+1}^n)\, d\pi_{n'}^{x_1^k, y_1^k}(x_{k+1}^{n'}, y_{k+1}^{n'}).
\end{aligned}$$

Since, by assumption, $\pi^{x_1^{n'}, y_1^{n'}} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu^{x_1^{n'}}, \nu^{y_1^{n'}})$, integrating over $y_{n'+1}^n$ yields

$$\int 1_A(x_{k+1}^n)\, d\mu^{x_1^{n'}}(x_{n'+1}^n)\, d\pi_{n'}^{x_1^k, y_1^k}(x_{k+1}^{n'}, y_{k+1}^{n'})$$

Now, we observe that, since $\pi_{n'} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_{n'}, \nu_{n'})$, by Lemma 3.1, $\pi_n^{x_1^k, y_1^k} \in \mathrm{Cpl}(\mu^{x_1^k}, \mu^{y_1^k})$. Thus, we can integrate over $y_{k+1}^{n'}$ and get

$$\int 1_A(x_{k+1}^n)\, d\mu^{x_1^{n'}}(x_{n'+1}^n)\, d\mu_{n'}^{x_1^k}(x_{k+1}^{n'})$$
$$\int 1_A(x_{k+1}^n)\, d(\mu^{x_1^k})^{x_{k+1}^{n'}}(x_{n'+1}^n)\, d\mu_{n'}^{x_1^k}(x_{k+1}^{n'})$$
$$= \mu^{x_1^k}(x_{k+1}^n \in A).$$

Repeating this with the roles of x and y interchanged completes the proof. □

The following proposition, proves that Knothe-Rosenblatt couplings and optimal bicausal couplings are optimal in the last time step, and we use this fact to bound the cost of the such couplings by the cost in the first $n-1$ time steps and the cost in the last time step.

Proposition 3.3. *For $p \geq 1$ and $\pi \in \mathrm{Cpl}_{\mathrm{KR}p}(\mu, \nu)$ or π a cost minimizing element in $\mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$ we have*

$$\left(\int \rho_{\mathcal{X}^n,p}^p \, d\pi\right)^{\frac{1}{p}} \leq \left(\int \sum_{i=1}^{n-1} \rho_{\mathcal{X}}(x_i, y_i)^p \, d\pi(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}} + \left(\int W_p(\mu^{\boldsymbol{x}}, \nu^{\boldsymbol{y}})^p \, d\pi_{n-1}(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}}.$$

Proof. First, note that

$$\int \rho_{\mathcal{X}}(x_n, y_n)^p \, d\pi(\boldsymbol{x}, \boldsymbol{y}) = \int \int \rho_{\mathcal{X}}(x_n, y_n)^p \, d\pi^{x_1^{n-1}, y_1^{n-1}}(x_n, y_n)\, d\pi_{n-1}(x_1^{n-1}, y_1^{n-1}).$$

If $\pi \in \mathrm{Cpl}_{\mathrm{KR}p}(\mu, \nu)$ we know that $\pi^{x_1^{n-1}, y_1^{n-1}} \in \mathrm{Cpl}(\mu^{x_1^{n-1}}, \nu^{y_1^{n-1}})$ is optimal for almost all $x_1^{n-1}, y_1^{n-1} \in \mathcal{X}^{n-1}$ and therefore $\int \rho_{\mathcal{X}}(x_n, y_n)^p \, d\pi^{x_1^{n-1}, y_1^{n-1}}(x_n, y_n) = W_p(\mu^{x_1^{n-1}}, \nu^{y_1^{n-1}})^p$ for almost all $x_1^{n-1}, y_1^{n-1} \in \mathcal{X}^{n-1}$. Thus

$$\left(\int \rho_{\mathcal{X}^n,p}^p \, d\pi\right)^{\frac{1}{p}} = \left(\int \sum_{i=1}^{n} \rho_{\mathcal{X}}(x_i, y_i)^p \, d\pi(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}}$$
$$= \left(\int \sum_{i=1}^{n-1} \rho_{\mathcal{X}}(x_i, y_i)^p \, d\pi(\boldsymbol{x}, \boldsymbol{y}) + \int \rho_{\mathcal{X}}(x_n, y_n)^p \, d\pi(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}}$$
$$= \left(\int \sum_{i=1}^{n-1} \rho_{\mathcal{X}}(x_i, y_i)^p \, d\pi(\boldsymbol{x}, \boldsymbol{y}) + \int W_p(\mu^{\boldsymbol{x}}, \nu^{\boldsymbol{y}})^p \, d\pi_{n-1}(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}}$$

The statement now follows from $x \to x^{\frac{1}{p}}$ being subadditive for $p \geq 1$.
If $\pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$ and minimizes cost, then by Lemma 3.1 $\pi^{x_1^{n-1}, y_1^{n-1}} \in \mathrm{Cpl}(\mu^{x_1^{n-1}}, \nu^{y_1^{n-1}})$ and $\pi^{x_1^{n-1}, y_1^{n-1}}$ must also be optimal for almost all $x_1^{n-1}, y_1^{n-1} \in \mathcal{X}^{n-1}$. Otherwise a measure $\bar{\pi}$

with $\bar{\pi}_n = \pi_n$ and $\bar{\pi}^{x_1^{n-1}, y_1^{n-1}} \in \mathrm{Cpl}(\mu^{x_1^{n-1}}, \nu^{y_1^{n-1}})$ optimal, would exist and have a lower cost then π and would also be in $\mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$ by Lemma 3.2 which is a contradiction. $\square$

Theorem 3.4. *If $({}^k\nu)_k$, a sequence of probability measures on $(\mathcal{X}^n, \rho_{\mathcal{X}^n})$, converges to a probability measure ν on $(\mathcal{X}^n, \rho_{\mathcal{X}^n,p})$ in the adapted p-Wasserstein distance then it converges also in the p-Knothe-Rosenblatt distance. In particular the p-Knothe-Rosenblatt distance induces the p-weak adapted topology.*

Proof. The proof is by induction. For $n = 1$ the statement follows because both, the Knothe-Rosenblatt and the adapted Wasserstein distance, coincide with the Wasserstein distance in this case.

$n \to n + 1:$

Let $({}^k\nu)_k$ be a sequence of probability measures on $(\mathcal{X}^{n+1}, \rho_{\mathcal{X}^{n+1},p})$ such that ${}^k\nu$ converges to a probability measure ν in the adapted Wasserstein distance. For arbitrary, fixed $({}^k\pi)_k$ with ${}^k\pi \in \mathrm{Cpl}_{\mathrm{KR}p}(\nu, {}^k\nu)$, we will show that the cost $\left(\int \rho^p_{\mathcal{X}^{n+1},p} \, d^k\pi\right)^{\frac{1}{p}}$ converges to 0 for $k \to \infty$.

By Proposition 3.3, the cost can be bounded by the cost in the first n dimension and the cost in the last dimension

$$\left(\int \rho^p_{\mathcal{X}^{n+1},p} \, d^k\pi\right)^{\frac{1}{p}} \leq \left(\int \sum_{i=1}^{n} \rho_{\mathcal{X}}(x_i, y_i)^p \, d^k\pi(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}} + \left(\int W_p(\nu^{\boldsymbol{x}}, {}^k\nu^{\boldsymbol{y}})^p \, d^k\pi_n(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}}.$$

Since $\left(\int \sum_{i=1}^n \rho_{\mathcal{X}}(x_i, y_i)^p \, d^k\pi(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}}$ coincides with the cost of the coupling $\pi_n \in \mathrm{Cpl}_{\mathrm{KR}p}({}^k\nu_n, \nu_n)$, it is bounded by $KR_p({}^k\nu_n, \nu_n)$. Now let $({}^k\gamma)_k$ be a sequence of bicausal couplings such that ${}^k\gamma$ is a cost minimizing element in $\mathrm{Cpl}_{\mathrm{BC}}({}^k\nu, \nu)$. Applying the triangle inequality for the p-Wasserstein distance in the second term yields

$$\left(\int W_p(\nu^{\boldsymbol{x}}, {}^k\nu^{\boldsymbol{y}})^p \, d^k\pi_n(\boldsymbol{x}, \boldsymbol{y})\right)^{\frac{1}{p}} \leq \left(\int (W_p(\nu^{\boldsymbol{z}}, {}^k\nu^{\boldsymbol{y}}) + W_p(\nu^{\boldsymbol{x}}, \nu^{\boldsymbol{z}}))^p \, d^k\pi_n \underset{\mathcal{Y}}{\otimes} {}^k\gamma_n(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\right)^{\frac{1}{p}}.$$

Using the Minkowski inequality and the definition of the product measure, the right hand side is bounded by

$$\left(\int W_p(\nu^{\boldsymbol{z}}, {}^k\nu^{\boldsymbol{y}})^p \, d^k\gamma_n(\boldsymbol{y}, \boldsymbol{z})\right)^{\frac{1}{p}} + \left(\int W_p(\nu^{\boldsymbol{x}}, \nu^{\boldsymbol{z}})^p \, d^k\pi_n \circledcirc {}^k\gamma_n(\boldsymbol{x}, \boldsymbol{z})\right)^{\frac{1}{p}}.$$

By Proposition 3.3 we get that $\left(\int W_p(\nu^{\boldsymbol{z}}, {}^k\nu^{\boldsymbol{y}})^p \, d^k\gamma_n(\boldsymbol{y},\boldsymbol{z})\right)^{\frac{1}{p}} \le AW_p({}^k\nu,\nu)$. To bound the second term we use the modulus of continuity and get

$$\left(\int W_p(\nu^{\boldsymbol{x}}, \nu^{\boldsymbol{z}})^p \, d^k\pi_n \circledcirc {}^k\gamma_n(\boldsymbol{x},\boldsymbol{z})\right)^{\frac{1}{p}} \le \omega_{\mathcal{I}(\nu),p}\left(\left(\int \sum_{i=1}^n \rho_{\mathcal{X}}(x_i,z_i)^p \; d^k\pi_n \circledcirc {}^k\gamma_n(\boldsymbol{x},\boldsymbol{z})\right)^{\frac{1}{p}}\right).$$

Using the triangle inequality for $\rho_{\mathcal{X}}$, the Minkowski inequality and that the modulus of continuity is increasing this can again be bounded by

$$\omega_{\mathcal{I}(\nu),p}\left(\left(\int \sum_{i=1}^n \rho_{\mathcal{X}}(x_i,y_i)^p \, d^k\pi(\boldsymbol{x},\boldsymbol{y})\right)^{\frac{1}{p}} + \left(\int \sum_{i=1}^n \rho_{\mathcal{X}}(y_i,z_i)^p \, d^k\gamma(\boldsymbol{y},\boldsymbol{z})\right)^{\frac{1}{p}}\right).$$

Applying Proposition 3.3 to the second part we see that this term is again bounded by $\omega_{\mathcal{I}(\nu),p}\left(KR_p({}^k\nu_n,\nu_n) + AW_p({}^k\nu,\nu)\right)$. Combining all the above inequalities yields

$$\left(\int \rho^p_{\mathcal{X}^{n+1},p} \, d^k\pi\right)^{\frac{1}{p}} \le KR_p({}^k\nu_n,\nu_n) + AW_p({}^k\nu,\nu) + \omega_{\mathcal{I}(\nu),p}\left(KR_p({}^k\nu_n,\nu_n) + AW_p({}^k\nu,\nu)\right) \tag{3.3}$$

Because $AW_p({}^k\nu,\nu) \to 0$ implies that $AW_p({}^k\nu_n,\nu_n) \to 0$ the induction hypothesis yields that also $KR_p({}^k\nu_n,\nu_n) \to 0$. Finally, since $\mathcal{I}(\nu)$ is concentrated on a graph of a function and the modulus of continuity is continuous at 0 [10, Lemma 2.7], we have $\omega_{\mathcal{I}(\nu),p}\left(KR_p({}^k\nu_n,\nu_n) + AW_p({}^k\nu,\nu)\right) \to 0$, because, by above argument $KR_p({}^k\nu_n,\nu_n) + AW_p({}^k\nu,\nu) \to 0$.

Since the right hand side in (3.3) is independent of the chosen sequence $({}^k\pi)_k$ we may take the supremum over all such sequences and obtain that $KR_p({}^k\nu,\nu) := \sup_{\pi\in\mathrm{Cpl}_{\mathrm{KR}_p}({}^k\nu,\nu)}\left(\int \rho^p_{\mathcal{X}^n,p} \, d\pi\right)^{\frac{1}{p}} \to 0$ for $k\to\infty$. ☐

4. The triangle inequality for the Knothe-Rosenblatt distance

We want to examine further properties of the Knothe-Rosenblatt distance. It is positive definite and symmetric but in general it does not satisfy the triangle inequality for dimension $d > 1$.

Lemma 4.1. *The p-Knothe-Rosenblatt distance on $\mathcal{P}((\mathbb{R}^d)^2, \|\,.\,\|_p)$ with $d > 1$ and $p \ge 1$ does not satisfy the triangle inequality.*

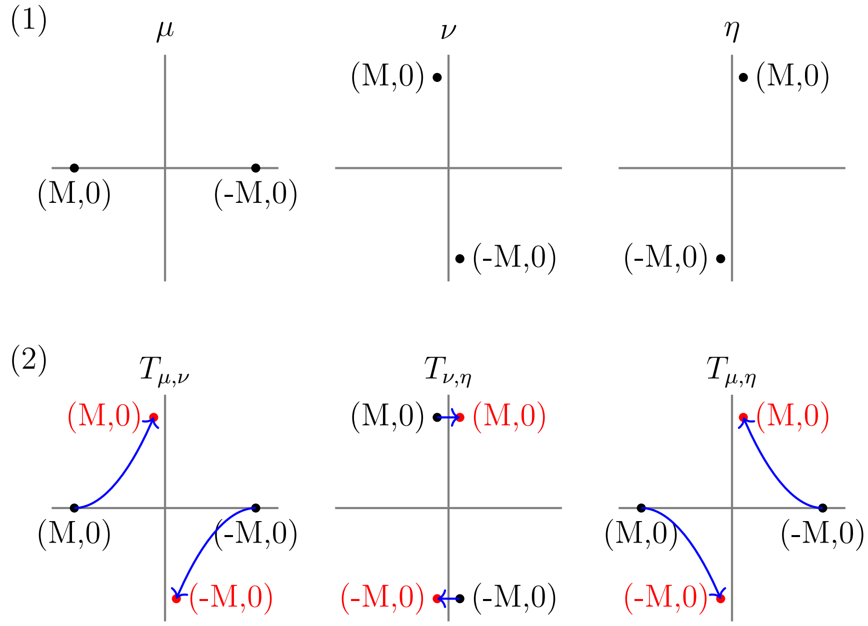

FIGURE 1. Example that the triangle inequality does not hold for the Knothe-Rosenblatt distance, see the proof of Lemma 4.1.

Proof. We will prove this by giving a counterexample for $d = 2$ for $d > 2$ one can easily find analogous counterexamples. Let $M \in \mathbb{R}$ be very large and define μ, ν, η probability measures on $((\mathbb{R}^2)^2, \|\,.\,\|_p)$ as

$$\begin{aligned}\mu =& \frac{1}{2}\left(\delta_{((-1,0),(M,0))} + \delta_{((1,0),(-M,0))}\right)\\ \nu =& \frac{1}{2}\left(\delta_{((-\varepsilon,1),(M,0))} + \delta_{((\varepsilon,-1),(-M,0))}\right)\\ \eta =& \frac{1}{2}\left(\delta_{((\varepsilon,1),(M,0))} + \delta_{((-\varepsilon,-1),(-M,0))}\right).\end{aligned}$$

In Figure 1(1) by plotting for each point with positive mass the first set of coordinates and labeling the points with the second set of coordinates. Between any of these measures there exits exactly one unique p-Knothe-Rosenblatt coupling. These couplings are also Monge and illustrated in Figure 1(2), we write $T_{\mu\nu}$ for the map from μ to ν. The red arrows indicate how the mass is moved, note that the Knothe-Rosenblatt coupling maps the first dimension optimally and thus only tries to minimize the movement of the points in the plane, ignoring the labels. The transport in the second dimension is predetermined by the transport in the plane since the conditional measures are all Dirac. Since

$$T_{\mu\eta}((-1,0),(M,0)) = ((-\varepsilon,-1),(-M,0)),$$
$$T_{\mu\eta}((1,0),(-M,0)) = ((\varepsilon,1),(M,0))$$

we have $KR_p(\mu,\eta) = (|1-\varepsilon|^p + |2M|^p + 1)^{\frac{1}{p}} \geq 2M$. Furthermore,

$$T_{\mu\nu}((-1,0),(M,0)) = ((-\varepsilon,1),(M,0)),$$
$$T_{\mu\nu}((1,0),(-M,0)) = ((\varepsilon,-1),(-M,0))$$

implies that $KR_p(\mu,\nu) = ((1-\varepsilon)^p + 1)^{\frac{1}{p}}$ and

$$T_{\nu\eta}((-\varepsilon,1),(M,0)) = ((\varepsilon,1),(M,0)),$$
$$T_{\nu\eta}((\varepsilon,-1),(-M,0)) = (-\varepsilon,-1),(-M,0))$$

implies that $KR_p(\nu,\eta) = 2\varepsilon$. Thus, for $0 < \varepsilon < 1 \ll M$ we have

$$KR_p(\mu,\nu) + KR_p(\nu,\eta) = ((1-\varepsilon)^p + 1)^{\frac{1}{p}} + 2\varepsilon < 2M < KR_p(\mu,\eta).$$

$\square$

Thus, because the triangle inequality does not hold, the p-Knothe-Rosenblatt distance is in general not a distance on $\mathcal{P}((\mathbb{R}^d)^2, \|\,.\,\|_p)$ with $d > 1$.

On the other hand, the p-Knothe-Rosenblatt distance is in fact a distance, satisfying the triangle inequality on $\mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$, i.e. the stochastic processes with $\mathbb{R}$ as the state space. To prove this, we will use another, but on $\mathcal{P}(\mathbb{R}^n)$ equivalent way of characterising Knothe-Rosenblatt couplings.

Definition 4.2. [25, Definition 2.3.5] *A probability measure $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ is a (strictly) increasing triangular transport if the support of π is (strictly) lexicographically monotone. A set $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is called lexicographically monotone if for $(x,y),(\widetilde{x},\widetilde{y}) \in \Gamma$ we have $x < \widetilde{x}$ implies $y \leq \widetilde{y}$ with regard to the lexicographical order on $\mathbb{R}^n$. The set Γ is called strictly lexicographically monotone if $x < \widetilde{x}$ implies $y < \widetilde{y}$.*

Definition 4.3. *For $\mu,\nu \in \mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$ we define $kr_{\mu\nu} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu,\nu)$ as the by* [25, Proposition 2.3.6] *unique triangular increasing bicausal transport from μ to ν. Furthermore we define*

Q^μ as the Monge transport plan that induces the coupling $kr_{\lambda^n\mu}$, where λ^n is the Lebesgue measure on $[0,1]^n$.

This $kr_{\mu\nu}$ is a Knothe-Rosenblatt coupling and if $KR(\mu,\nu)$ is finite then this the only Knothe-Rosenblatt coupling from μ to ν. In Proposition 4.6 we will show that $kr_{\mu\nu} = (Q^\mu, Q^\mu)_\# \lambda^n$ and from this statement the triangle inequality will follow immediately using the Minkowski inequality. This approach is similar to the one used in [7], where the Knothe-Rosenblatt coupling is defined by $(Q^\mu, Q^\mu)_\# \lambda^n$ and this definition is used to extend the Knothe-Rosenblatt distance to $\mathcal{P}((\mathbb{R}^d)^n, \|\,.\,\|_p)$ in a way that satisfies the triangle inequality and induces the adapted Wasserstein topology.

To prove Proposition 4.6 we first show in two auxiliary lemmas that the marriage of two bicausal couplings is bicausal and that the marriage of a strictly triangular increasing and a triangular increasing coupling is triangular increasing.

Lemma 4.4 (The marriage of two bicausal transport plans is bicausal)**.** *Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ and $\pi \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_1,\mu_2), \gamma \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_2,\mu_3)$. Then $\pi \varpi \gamma \in \mathrm{Cpl}_{\mathrm{BC}}(\mu_1,\mu_3)$.*

Proof. Let (X,Y,Z) be the random vector with distribution $\pi \underset{\mathcal{Y}}{\otimes} \gamma$. We have to show that for all $k \le n$ the map $x_1^n \to (\pi\varpi\gamma)^{x_1^n}(Z_1^k \in B)$ is $\mathcal{F}_{X_1^k}$ measurable for all measurable $B \subseteq \mathcal{X}^k$, where $(\mathcal{F}_{X_1^k})_{k\ge 1}$ is the natural filtration of X. Note that $(\pi\varpi\gamma)^x(B) = \int_{\mathcal{X}^n} \gamma^y(Z_1^k \in B) d\pi^x(y)$, and due to γ and π being causal, $y \to \gamma^y(Z_1^k \in B)$ is $\mathcal{F}_{Y_1^k}$ measurable, non-negative and bounded, and $x \to \pi^x(Y_1^k \in B)$ is $\mathcal{F}_{X_1^k}$ measurable. Thus the composition is $\mathcal{F}_{X_1^k}$ measurable and hence $\pi\varpi\gamma$ is causal. Exchanging the roles of X and Z above, yields that $\pi\varpi\gamma$ is bicausal. □

Lemma 4.5 (The marriage of a strictly triangular increasing and a triangular increasing coupling is triangular increasing)**.** *Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$, $\pi \in \mathrm{Cpl}(\mu_1,\mu_2)$ strictly triangular increasing, and $\gamma \in \mathrm{Cpl}(\mu_2,\mu_3)$ triangular increasing. Then $\pi \varpi \gamma \in \mathrm{Cpl}(\mu_1,\mu_3)$ is triangular increasing.*

Proof. Let $(x,z), (\widetilde{x},\widetilde{z}) \in supp(\pi\varpi\gamma)$ then there exist $y, \widetilde{y} \in \mathbb{R}^n$ such that $(x,y), (\widetilde{x},\widetilde{y}) \in supp(\pi)$ and $(y,z), (\widetilde{y},\widetilde{z}) \in supp(\gamma)$. Thus, if $x < \widetilde{x}$ we have, since π is strictly triangular increasing, that $y < \widetilde{y}$. Now by γ being triangular increasing it follows that $z \le \widetilde{z}$. □

Proposition 4.6. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$. Then*

$$kr_{\mu,\nu} = (Q^\mu, Q^\nu)_\# \lambda^n.$$

Proof. To prove this statement it suffices to show that $(Q^\mu, Q^\nu)_\# \lambda^n \in \mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$ and that it is triangular increasing. To see this, we observe that

$$(Q^\mu, Q^\nu)_\# \lambda^n = (Q^\mu, id)_\# \lambda^n \circledcirc (id, Q^\nu)_\# \lambda^n.$$

Now Lemma 4.4 gives us $(Q^\mu, Q^\nu)_\# \lambda^n \in \mathrm{Cpl}_{\mathrm{BC}}(\mu, \nu)$. This follows since $(Q^\mu, id)_\# \lambda^n = kr_{\mu\lambda^n}$ and $(id, Q^\nu)_\# \lambda^n = kr_{\lambda^n \nu}$ are both bicausal.
Furthermore, since $supp(kr_{\lambda^n \mu}) = \{(y, Q^\nu(y)) | y \in [0,1]^n\}$ is lexicographically monotone,

$$supp((Q^\nu, id)_\# \lambda^n) = \{(Q^\nu(y), y) | y \in [0,1]^n\} = \{(y, x) | (x, y) \in supp(kr_{\lambda^n \mu})\}$$

is strictly lexicographically monotone, and thus, $(Q^\nu, id)_\# \lambda^n$ is strictly triangular increasing. Also, by definition, $(id, Q^\nu)_\# \lambda^n = kr_{\lambda^n \nu}$ is triangular increasing. Thus, by Lemma 4.5 also $(Q^\mu, Q^\nu)_\# \lambda^n$ is triangular increasing and hence $kr_{\mu,\nu} = (Q^\mu, Q^\nu)_\# \lambda^n$. □

Proposition 4.7. *Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(\mathbb{R}^n, \|\,.\,\|_p)$, then*

$$KR_p(\mu_1, \mu_3) \leq KR_p(\mu_1, \mu_2) + KR_p(\mu_2, \mu_3).$$

Proof. By Proposition 4.6 we have

$$\begin{aligned}
KR_p(\mu_1, \mu_3) &= \left(\int \|x - z\|_p^p \, d(Q^{\mu_1}, Q^{\mu_3})_\# \lambda^n(x, z) \right)^{1/p} \\
&\leq \left(\int (\|x - y\|_p + \|y - z\|_p)^p \; d(Q^{\mu_1}, Q^{\mu_2}, Q^{\mu_3})_\# \lambda^n(x, y, z) \right)^{1/p} \qquad (4.1) \\
&\leq \left(\int \|x - y\|_p^p \, d(Q^{\mu_1}, Q^{\mu_2})_\# \lambda^n(x, y) \right)^{1/p} + \left(\int \|y - z\|_p^p \, d(Q^{\mu_2}, Q^{\mu_3})_\# \lambda^n(x, y) \right)^{1/p} \qquad (4.2) \\
&= KR_p(\mu_1, \mu_2) + KR_p(\mu_2, \mu_3).
\end{aligned}$$

Here we used in (4.1) the triangle inequality for the p-norm on $\mathbb{R}^n$ and in (4.2) the Minkowski inequality. □

5. Precompactness

Let $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ be the set of stochastic processes on a compact set $\mathcal{X}$ with n time steps. In this chapter, we show that while $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ equipped with the adapted Wasserstein distance is precompact (Lemma 5.4), the same is not necessarily true for $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ equipped with the Knothe-Rosenblatt distance (Prop 5.2). We will start by proving that for $d \geq 1$, $\mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$, equipped with the p-Knothe-Rosenblatt distance, is not precompact by giving an example of a sequence without a convergent subsequence. But since we showed that the Knothe-Rosenblatt distance in general does not satisfy the triangle inequality, this does not immediately yield precompactness and we need the following lemma.

Lemma 5.1. *Let $\nu \in \mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ and $d \geq 1$. Then there exist an almost surely unique $\tilde{\nu} \in \mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ such that $\tilde{\nu}_1 = \lambda_{[0,1]^d}$ and for any $\mu \in \mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ with $\mu_1 = \lambda_{[0,1]^d}$ we have that $KR_p(\nu, \mu) \geq KR_p(\tilde{\nu}, \mu)$ for any $p \geq 1$.*

Proof. Since $\lambda_{[0,1]^d}$ is absolutely continuous, there exists an optimal transport plan γ between ν_1 and $\lambda_{[0,1]^d}$. Since $\lambda_{[0,1]^d}$ is absolutely continuous, γ is almost surely unique by Brenier's Theorem [24][Theorem 4.2]. We now choose π as the coupling such that $\pi_1 = \gamma$ and $\pi^{x,y} = (x \to (x,x))_{\#}\nu^x$ and set $\tilde{\nu} = \mathrm{proj}_2\pi$. Let $\mu \in \mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ with $\mu_1 = \lambda_{[0,1]^d}$ then by Prop. 3.3 we have that

$$KR_p(\nu, \mu) \geq \left(\int_{([0,1]^d)^2} W(\nu^x, \mu^y)^p \, d\gamma(x,y) \right)^{\frac{1}{p}}.$$

By construction,

$$\int_{([0,1]^d)^2} W(\nu^x, \mu^y) \, d\gamma(x,y) = \int_{[0,1]^d} W(\tilde{\nu}^x, \mu^x) \, dx$$

and since $\tilde{\nu}_1 = \mu_1 = \lambda_{[0,1]^d}$, using Prop. 3.3,

$$\int_{[0,1]^d} W_p(\tilde{\nu}^x, \mu^x) \, dx = KR_p(\tilde{\nu}, \mu).$$

$\square$

Proposition 5.2. *The set of $\mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ equipped with the p-Knothe-Rosenblatt distance is not precompact. Indeed, no finite set of open balls of size $2^{-(p+1)}$ covers $\mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$.*

Proof. Let $f_n : [0,1] \to \{0,1\}$ such that $f_n(x) = \lfloor 2^n x \rfloor - 2\lfloor 2^{n-1}x \rfloor$ and ${}^n\mu = (id, f_n \circ \mathrm{proj}_1)_\# \lambda_{[0,1]^d}$ where $\lambda_{[0,1]^d}$ is the Lebesgue measure on $[0,1]^d$. For a visual representation of ${}^1\mu, {}^2\mu, {}^3\mu$ see Figure 2, it shows how the distribution in the second variable (or time step) depends on the first. The distribution of the second variable conditional on the first variable starting in the black area is the Dirac measure δ_1. Analogously, conditional on the first variable starting in the red area the distribution of the second variable is δ_0. By construction ${}^n\mu_1 = \lambda_{[0,1]^d}$ for all n. Thus for $n \neq m$ we have

$$KR_p({}^n\mu, {}^m\mu) = \left(\int_0^1 |f_n(x) - f_m(x)|^p \right)^{\frac{1}{p}} dx = 2^{-p}.$$

Note that since by Lemma 4.1, for $d > 1$, the Knothe-Rosenblatt distance on $\mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ does not satisfy the triangle inequality, the existence of this sequence is not immediately equivalent to precompactness.

Assume that there exist ${}^1v, \dots, {}^l\nu \in \mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$ such that $\cup_{i=1}^l B_\varepsilon({}^i\nu)$ is an ε-covering of $\mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$. Then by Lemma 5.1 there exist ${}^1\tilde{\nu}, \dots, {}^l\tilde{\nu}$ such that ${}^i\tilde{\nu}_1 = \lambda_{[0,1]^d}$ and for any i,j that $KR_p({}^i\nu, {}^j\mu) \geq KR_p({}^i\tilde{\nu}, {}^j\mu)$. Let $i \in [1, \dots, l]$ and $n \neq m$ and $KR_p({}^i\nu, {}^m\mu) < \varepsilon < 2^{-(p+1)}$. Then, since ${}^n\mu_1 = {}^n\nu_1 = {}^i\tilde{\nu}_1 = \lambda_{[1,0]^d}$ we have by Prop. 3.3 that

$$\begin{aligned}
KR_p({}^i\nu, {}^n\mu) \geq KR_p({}^i\tilde{\nu}, {}^n\mu) = & \left(\int_{[0,1]^d} W_p({}^i\tilde{\nu}^x, {}^n\mu^x)^p \, dx \right)^{\frac{1}{p}} \\
\geq & \left(\int_{[0,1]^d} \left(W_p({}^m\mu^x, {}^n\mu^x) - W_p({}^i\tilde{\nu}^x, {}^m\mu^x) \right)^p \, dx \right)^{\frac{1}{p}} && (5.1) \\
\geq & \left(\int_{[0,1]^d} W_p({}^m\mu^x, {}^n\mu^x)^p \, dx \right)^{\frac{1}{p}} - \left(\int_{[0,1]^d} W_p({}^i\tilde{\nu}^x, {}^m\mu^x)^p \, dx \right)^{\frac{1}{p}} && (5.2) \\
= & KR_p({}^m\mu, {}^n\mu) - KR_p({}^i\tilde{\nu}, {}^m\mu) \geq 2^{-p} - \varepsilon > \varepsilon.
\end{aligned}$$

In (5.1) we have applied the reverse triangle inequality for the p-Wasserstein distance distance and in (5.2) the reverse triangle inequality for the L_p-norm.

Therefore, each $B_\varepsilon(\nu^j)$ can contain at most one element of $({}^i\mu_i)_{i \geq 1}$, and thus $\cup_{i=1}^l B_\varepsilon({}^i\nu)$ cannot be an ε-covering of $\mathcal{P}(([0,1]^d)^2, \|\,.\,\|_p)$. $\square$

Remark 5.3. *Using similar methods proof can be be extended to show that for any $n \geq 2$ it holds that $\mathcal{P}(([0,1]^d)^n, \|\,.\,\|_p)$ equipped with the p-Knothe-Rosenblatt distance is not precompact.*

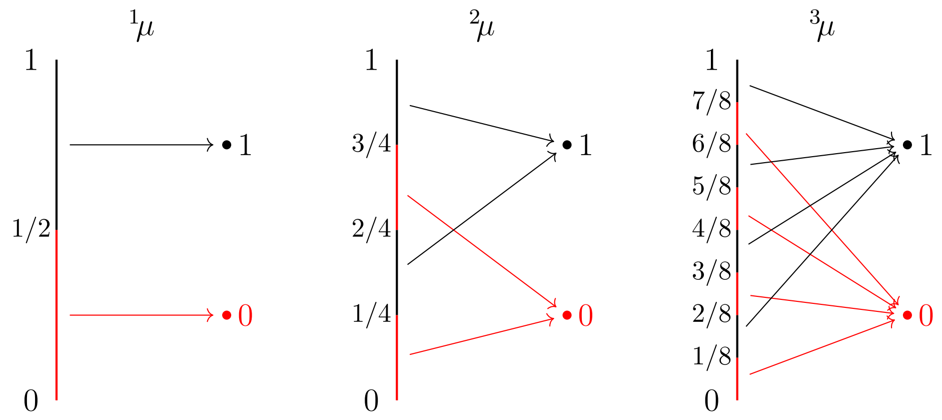

FIGURE 2. Visual representation of the measures $^1\mu, ^2\mu, ^3\mu$ from Proposition 5.2. The figure shows how the distribution in the second variable (or time step) depends on the first. In the first time step we have a uniform distribution symbolised be the red and black line. The colours and arrows indicate whether in the second time step the stochastic process will be 0 or 1. I.e. the distribution of the second variable conditional on the first variable starting in the black area is the Dirac measure δ_1. Analogously, conditional on the first variable starting in the red area the distribution of the second variable is δ_0.

In contrast, equipped with the adapted Wasserstein distance, $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ is indeed precompact, for compact spaces $\mathcal{X}$. We will prove this by induction over n, by showing that $\mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^{n+1})$ is isomorphic to a subset of $\mathcal{P}_{\mathcal{W}_p}(X \times \overline{\mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^n)})$.

Lemma 5.4. *Let $\mathcal{X}$ be a compact set. Then $\mathcal{P}(\mathcal{X}^n, \|\,.\,\|_p)$ equipped with the adapted Wasserstein distance is precompact.*

Proof. For this proof we use the following notation. For a Polish metric space A, we refer to the space of probability measures on A, equipped with the p-Wasserstein distance, by $\mathcal{P}_{\mathcal{W}_p}(A)$ and the space of probability measures on A, equipped with the adapted p-Wasserstein distance, by $\mathcal{P}_{\mathcal{AW}_p}(A)$.

Proof by induction over n. For $n = 1$ the adapted p-Wasserstein distance coincides with the normal p-Wasserstein distance. Hence the statement follows from $\mathcal{P}_{\mathcal{W}_p}(\mathcal{X})$ being compact with regard to the weak topology since $\mathcal{X}$ is compact.

For $n \to n+1$ we will show that

$$\mathcal{I}_1 : \mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^{n+1}) \to \mathcal{P}_{\mathcal{W}_p}(\mathcal{X} \times \mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^n))$$

with $\mathcal{I}_1(\nu) := ((x \to (x, \nu^x))_{\#}\nu_1$ is an isometry but not bijective.

$$\mathcal{AW}_p(\mu,\nu)^p = \inf_{\pi\in \mathrm{Cpl}_{\mathrm{BC}}(\mu,\nu)} \int \sum_{i=1}^{n+1} ||x_i - y_i||_p^p \, d\pi(\boldsymbol{x},\boldsymbol{y})$$
$$= \inf_{\pi\in \mathrm{Cpl}_{\mathrm{BC}}(\mu,\nu)} \int\int \sum_{i=2}^{n+1} ||x_i - y_i||_p^p \; d\pi^{x_1,y_1}(\boldsymbol{x}_2^{n+1},\boldsymbol{y}_2^{n+1}) + ||\boldsymbol{x}_1 - \boldsymbol{y}_1||_p^p \, d\pi_1(x_1,y_1)$$

By Lemma 3.2 we have $\pi^{x_1,y_1} \in \mathrm{Cpl}_{\mathrm{BC}}(\mu^{x_1},\nu^{y_1})$ almost everywhere. Furthermore, π^{x_1,y_1} is optimal in $\mathrm{Cpl}_{\mathrm{BC}}(\mu^{x_1},\nu^{y_1})$ π_1 almost everywhere. Otherwise, by Lemma 3.2, we could construct a bicausal coupling between μ and ν with a lower cost than π. Hence,

$$\int \sum_{i=2}^{n+1} ||x_i - y_i||_p^p \; d\pi^{x_1,y_1}(\boldsymbol{x}_2^{n+1},\boldsymbol{y}_2^{n+1}) = \mathcal{AW}_p(\mu^{x_1},\nu^{y_1})^p.$$

Therefore,

$$\mathcal{AW}_p(\mu,\nu)^p = \inf_{\pi_1\in \mathrm{Cpl}(\mu_1,\nu_1)} \int_{\mathcal{X}^2} ||x_1 - y_1||_p^p + \mathcal{AW}_p(\mu^{x_1},\nu^{y_1})^p \, d\pi_1(x_1,y_1).$$

Note, that we replaced $\mathrm{Cpl}_{\mathrm{BC}}(\mu_1,\nu_1)$ with $\mathrm{Cpl}(\mu_1,\nu_1)$, since all couplings between stochastic process with only one time points are bicausal.

Furthermore, since $\mathcal{I}_1(\mu)$ is concentrated on the graph of $x \to (x,\mu^x)$, this equals

$$\inf_{\widetilde{\pi}\in \mathrm{Cpl}(\mathcal{I}_1(\mu),\mathcal{I}_1(\nu))} \int_{(\mathcal{X}\times\mathcal{P}(\mathcal{X}^n))^2} ||x_1 - y_1||_p^p + \mathcal{AW}_p(\gamma,\varphi)^p \, d\widetilde{\pi}((x_1,\gamma),(y_1,\varphi)),$$

which equals $\mathcal{W}_p(\mathcal{I}_1(\mu),\mathcal{I}_1(\nu))^p$ by definition of the p-Wasserstein distance.

Precompactness now follows since $\mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^{n+1})$ is isomorphic to a subset of

$$\mathcal{P}_{\mathcal{W}_p}(X \times \mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^n)) \subseteq \mathcal{P}_{\mathcal{W}_p}(X \times \overline{\mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^n)})$$

and $\mathcal{P}(X \times \overline{\mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^n)})$ equipped with the weak topology is compact, since $\mathcal{X}$ and, by the induction hypothesis, also $\overline{\mathcal{P}_{\mathcal{AW}_p}(\mathcal{X}^n)}$ are compact. $\square$

References

[1] D. Aldous. Weak convergence and the theory of processes. *University of California, Berkeley*, 1981.

[2] J. Backhoff, M. Beiglbock, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

[3] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.*, 24(3):601–632, 2020.

[4] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probability Theory and Related Fields*, 178(3):1125–1172, 2020.

[5] D. Bartl and J. Wiesel. Sensitivity of multiperiod optimization problems with respect to the adapted wasserstein distance. *SIAM J. Financ. Math.*, 14(2):704–720, 2023.

[6] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Approximation of martingale couplings on the line in the adapted weak topology. *Probab. Theory Related Fields*, 183(1-2):359–413, 2022.

[7] M. Beiglböck, G. Pammer, and A. Posch. The Knothe-Rosenblatt distance and its induced topology. University of Vienna (preprint), 2023.

[8] J. Bion-Nadal and D. Talay. On a Wasserstein-type distance between solutions to stochastic differential equations. 2019.

[9] M. Cullen and H. Maroofi. The fully compressible semi-geostrophic system from meteorology. *Archive for rational mechanics and analysis*, 167(4):309–336, 2003.

[10] M. Eder. Compactness in adapted weak topologies. *arXiv preprint arXiv:1905.00856*, 2019.

[11] A. Genevay, M. Cuturi, G. Peyré, and F. Bach. Stochastic optimization for large-scale optimal transport. In *Advances in neural information processing systems*, pages 3440–3448, 2016.

[12] M. Glanzer, G. C. Pflug, and A. Pichler. Incorporating statistical model error into the calculation of acceptability prices of contingent claims. *Mathematical Programming*, 174:499–524, 2019.

[13] M. Glanzer, G. C. Pflug, and A. Pichler. Incorporating statistical model error into the calculation of acceptability prices of contingent claims. *Math. Program.*, 174(1-2, Ser. B):499–524, 2019.

[14] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of computer vision*, 60(3):225–240, 2004.

[15] M. F. Hellwig. Sequential decisions under uncertainty and the maximum theorem. *Journal of Mathematical Economics*, 25(4):443–464, 1996.

[16] K. B. Kirui, G. C. Pflug, and A. Pichler. New algorithms and fast implementations to approximate stochastic processes. 2020.

[17] H. Knothe. Contributions to the theory of convex bodies. *Mich. Math. J.*, 4:39–52, 1957.

[18] R. Lassalle. Causal transport plans and their Monge-Kantorovich problems. *Stochastic Analysis and Applications*, 36(3):452–484, 2018.

[19] G. Peyré, M. Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[20] P. Pfanzagl. Conditional distributions as derivatives. *The Annals of Probability*, 7(6):1046–1050, 1979.

[21] G. C. Pflug. Version-independence and nested distributions in multistage stochastic optimization. *SIAM Journal on Optimization*, 20(3):1406–1420, 2009.

[22] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.

[23] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23(3):470–472, 1952.

[24] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[25] A. Zalashko. *Causal optimal transport: theory and applications*. PhD thesis, PhD Thesis, University of Vienna, 2017.